

NyS

Titel:	Et forsøg på at explicere begrebet "Semantisk struktur"
Forfatter:	Peter Bøgh Andersen
Kilde:	<i>NyS – Nydanske Studier</i> & <i>Almen kommunikationsteori</i> 4, 1972, s. 36-63
Udgivet af:	Akademisk Forlag
URL:	www.nys.dk



© NyS og artiklens forfatter

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre NyS-numre (NyS 1-36) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

ET FORSØG PÅ AT EXPLICERE BEGREBET 'SEMANTISK STRUKTUR'.

Af Peter Bøgh Andersen

Aarhus Universitet

Semantikken er den disciplin der længst og mest hårdnakket har modsat sig en rimelig formalisering. Forsøg har naturligvis været gjort, men ingen tilfredsstillende. Katz og Fodor har æren af at have fremsat den første formaliserede teori; den er imidlertid blevet hårdt kritiseret (Weinreich (66)), og er efter min formening deskriptivt inadækvat, fordi den ganske negligerer sprogudøverens 'referentielle' viden, dvs. de regler der angår sætningernes reference.

Den formelle grammatiks rivende udvikling i disse år ligger vel tildels i, at man har gjort sig klart præcist hvilke fænomener den grammatiske beskrivelse skal gøre rede for. Måske beror semantikens dvaletilstand på, at man endnu ikke i denne disciplin har fastlagt hvilke af sprogudøverens evner de semantiske regler skal gøre rede for.

Måske skyldes vanskelighederne også at der ikke findes ét men mange overlappende semantiske systemer. Når en botaniker taler vil indholdet af hans sætninger givetvis indeholde betydningsstrukturer som stammer dels fra dagligsproget, dels fra botanikkens klassificeringssystem.

I det følgende skal jeg opridse en model, hvormed man kan beskrive visse sider af sprogets indhold. Selv har jeg hovedsageligt anvendt den til beskrivelse af sætninger, som betegner et objekts placering eller bevægelsesretning, og det må stå åbent, om den også kan finde videre anvendelse. Modellen er opstillet med særligt henblik på de problemer man træffer indenfor den generative grammatik.

Det gælder først om at fastslå, hvilke sproglige fænomener man vil anse for semantisk relevante. Jeg opstiller tre: 1) implikationsforhold mellem sætninger, 2) ordenes distributionsforhold og 3) forholdet mellem det sproglige udtryk og de oplevede ting, som udtrykket beskriver. Det sidste punkt forvises ofte fra lingvistikken (således i Katz/Fodor-teorien); men det er et faktum, at mennesket bruger sætninger til at beskrive ting med, samt at det også her følger bestemte regler. Referenceregler skal derfor også med i en eksplicit beskrivelse af sprogudøverens regelsystem.

Man finder ofte at sætninger kan implicere hinanden. Eksempelvis vil sætningerne

- (1) Jeg så flyveren styrte ned.
- (2) Den kloge mand stod ved vinduet.

implicere henholdsvis

- (3) Flyveren styrtede ned.
- (4) Manden er klog.

Når man hævder (1) hævder man samtidig (3) og når (2) hævdes hævdes (4) samtidig. Man kan da sige, at (3) og (4) er komponenter i (1) og (2). I transformationsgrammatikken forklarer man dette ved, at der under (1) ligger en abstrakt struktur, som netop indeholder (3) som en delstruktur. Den abstrakte struktur, sætningsbasen, kunne være:

- (5) (_S Jeg så (_S flyveren styrtede ned)_S)_S

Denne analyse er gængs også udenfor transformationsgrammatikken; Diderichsen taler her om sekundær nexus. For at få frembragt (1) af (5) formulerer man en transformationsregel: slet præteritumsmærket! Idet infinitiv opfattes som den ikke-mærkede form, får man netop (1) af (5). Transformationer består af slettelser, omflytninger eller tilføring af en eller anden konstant. Man plejer at regne med, at transformationer ikke må berøre sætningens mening; basestrukturen er derfor den eneste syntaktiske struktur, som er relevant, når man skal tilskrive sætningen et indhold. Derfor kan (6) ikke være transformeret af (7):

- (6) Kommer han?
- (7) Han kommer.

Man postulerer da et spørgemorfem i sætningsbasen, som ligger under (6). Dette spørgemorfem slettes, verbum og subjekt byttes rundt og sætningen får stigende intonation, hvorved (6) fremkommer.

Betragt nu følgende sætninger:

- (8) Det er ikke Oles bold.

(9) Hans lånte bolden til Ole.

(10) Ole har bolden nu.

(11) Ole har taget bolden.

Man ser let, at ligesom (1) før implicerede (3), således implicerer nu (9) (8) og (11) (10); det er nemlig klart, at man vil foretrække det øverste af følgende alternativer:

(12)a Hans lånte bolden til Ole, så det er ikke Oles bold.

b Hans lånte bolden til Ole, så det er Oles bold.

(13)a Ole har taget bolden, så han har bolden nu.

b Ole har taget bolden, så han har ikke bolden nu.

Det ser da ud til, at man, når man hævder (9) og (11), samtidig hævder henholdsvis (8) og (10); man skulle tro, at (8) måtte anbringes i basestrukturen til (9). Dette kan imidlertid ikke lade sig gøre; der er nemlig en fundamental forskel mellem de to tilfælde. I første tilfælde var det muligt at frembringe de transformerede sætninger af deres basestruktur ved hjælp af regler, som slettede eller byttede rundt på led. Disse regler er generelle og fungerer uden hensyn til hvilke ord der står på verbernes og substantivernes plads; en basestruktur som

(14) (_S Jeg så (_S Hans lånte bolden til Ole)_S)_S

vil konverteres på akkurat samme måde som (5), nemlig til

(15) Jeg så Hans låne bogen til Ole.

skønt ordene i den indkapslede sætning er forskellige i (5) og (15). Forholdet er anderledes i andet tilfælde. Her beror implikationsforholdene på de faktisk forekommende ord; skifter man ordene ud, ophører implikationerne. Man vil ikke umiddelbart foretrække et af følgende alternativer:

(16)a Hans kastede bolden til Ole, så det er ikke Oles bold.

b Hans kastede bolden til Ole, så det er Oles bold.

Det viser, at sætningen "Hans kastede bolden" ikke implicerer "Det er ikke Oles bold". Eftersom disse relationer øjensynligt beror på

de enkelte ord, er "idiosynkratiske" for det enkelte ordpar, så er det ikke muligt at anvende transformationer her; transformationer opererer jo ikke på det enkelte ord, men på syntaktiske symboler (de refererer til "subjekt", "verbum", osv.), og transformationsreglerne skal være generelle. Nu er netop betydningen idiosynkratisk for det enkelte ord. Det ligger da nær at forklare implikationsforholdene som funktionelt betingede af ordenes mening. Man er åbenbart nødt til at postulere et system, som kan kaldes semantisk og som "går på tværs" af de syntaktiske regler. Dette system kan jævnføres med glossematikens "indholdsform". Jeg kalder det "indholdssystemet". Det gælder nu om at formulere kodningsregler mellem udtryk og indhold.

En præcis beskrivelse af kodningsproceduren skal jeg først senere komme ind på. På dette indledende trin skal jeg formulere forholdet så simpelt som muligt. Der opstilles en relation T . T består af en samling af ordnede par $(x_1, y_1), \dots, (x_n, y_n)$, hvor x 'erne er udtryksstørrelser og y 'erne er indholdsstørrelser. Det er lidet sandsynligt at T er en funktion i ordets matematiske bemærkelse. Derfor kan Hjeltslevs term "tegnfunktion" være vildledende.

Sætning 11 har sætning 10 som komponent. Den har også følgende sætning som komponent:

(17) Ole havde ikke bolden før.

idet man vil foretrække det øverste af alternativerne

(18)a Ole har taget bolden, så han havde den ikke før.

b Ole har taget bolden, så han havde den før.

Åbenbart kan man opfatte betydningen af sætning 11 som to sætninger, nemlig sætning 10 og 17. Et element i relationen T er da følgende:

$(\{ \text{Ole har taget bolden} \}, \{ \text{Ole havde ikke bolden før, Ole har bolden nu} \})$

Indholdet af sætning 11 er da en mængde på to elementer, nemlig sætning 10 og 17. Tuborg-parenteserne symboliserer en uordnet mængde. I Wallace og Atkins (60) foreslås det netop, at man skal

anvende sådanne uordnede mængder til beskrivelse af sprogets indholdsside. I visse tilfælde (slægtskabssemantik) synes det at kunne lade sig gøre, men generelt må man nok stille sig tvivlende; ofte er det praktisk at ordne de semantiske elementer og ikke blot notere dem som en uordnet mængde.

Sætning 11 indeholder en kausativskomponent: det er subjektets referent, som bringer den forandring i stand, som sætningen meddeler. Man kan overbevise sig om dette ved at prøve at parre sætning 19 og 20 sammen med sætning 21:

- (19) Hvad foretog Ole sig så?
- (20) Hvad skete der så med Ole?
- (21) Ole tog bolden.

(21) fungerer udmærket som svar på (19) men kan ikke være svar på (20). Derfor er det nødvendigt at finde en måde, hvorpå man kan repræsentere, at det er Ole som bevirker, at han fra ikke at have bolden kommer til at have den. Men dette kan - så vidt jeg kan se - ikke lade sig gøre i den notation, som hidtil er fulgt.

I stedet for at opfatte (10) og (17) som komponenter, kan man opfatte følgende sætninger som komponenter:

- (22) Ole har ikke bolden.
- (23) Ole har bolden.

Nu er det ikke længere ligegyldigt i hvilken rækkefølge de to komponentsætninger optræder. Den situation, som (22) betegner ("Ole har ikke bolden"), skal gå umiddelbart forud for den situation, som (23) refererer til ("Ole har bolden"). De to sætninger må være ordnet i tid:

- (24) (Ole har ikke bolden) og så (har han bolden).

I sætning 10 og 17 var denne ordning udtrykt ved adverbierne "før" og "nu". Øjensynligt kan indholdet af sætning 11 beskrives som to elementer imellem hvilke der hersker en tidsrelation. Indholdssystemet må derfor være et system, hvori man kan notere relationer; dette kan man i de logiske sprog, og man kan da gætte på, at et system, der minder om et logisk sprog, kan anvendes som refe-

rencesystem. Det er imidlertid væsentligt, at man opgiver en normerende holdning til sproget; når filosoffer foretager en logisk transskription af dagligsproget sker det ofte i kritisk øjemed for at påvise dårligheder i dagligsproget. Set fra en sproglig synsvinkel er denne fremgangsmåde uholdbar; derfor må det fastholdes, at de logiske tegnsprog kun må betragtes som referencesystemer og ikke som normgivende systemer.

Termen "sprog" bruges ofte dobbelttydig; dels om det, man almindeligvis kalder sprog, dels om kæder af symboler, som er bestemt ved en grammatik. Denne forvirring er gennemgående indenfor den generative grammatik (se f.eks. Hockett (66)). I det følgende vil termen "sprog" kun blive brugt i den almene betydning; om formelle systemer, karakteriseret ved en grammatik, vil Hockett's udtryk "harpe" blive anvendt.

Et sprogs udtryksside består af strenge af symboler, som kædes sammen. Ikke alle strenge kan sammenkædes. En grammatik udskiller de strenge, som er tilladte, fra de strenge, som ikke er det; den karakteriserer sproget ved ud af de uendelig mange kombinationer at vælge dem, som er tilladte i det pågældende sprog; heraf navnet "harpe" (en harpe består jo af "strenge"). En grammatik, som karakteriserer en harpe, siger noget om sprogets syntaks, men ikke noget om betydningssiden.

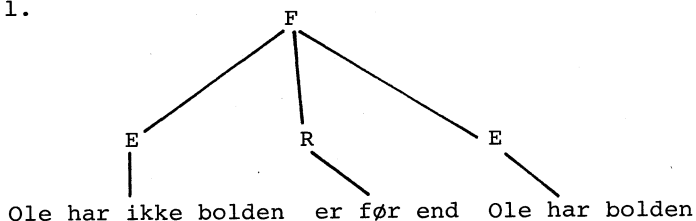
Man kan også opfatte de logiske tegnsprog som harper, dvs. som strenge af elementer, hvis kombination er styret af grammatiske regler (se f.eks. Chomsky (63)).

Hvis det er rigtigt, at indholdssystemet i sproget i visse henseender minder om de logiske sprog, så kunne man tænke sig, at sprogets indholdsside ligesom dets udtryksside kunne opfattes som en harpe, dvs. som strenge af symboler, der kan sammenkædes efter bestemte regler.

Indholdet af sætning 21 kunne opfattes som to sætninger, der var ordnet i tid. To elementer, som er ordnet i en vis relation til hinanden, kaldes nu et forhold F . Et forhold består af en relation R og elementerne, som ordnes. De kaldes E . Symbolerne F , R og E er semantiske kategorier på samme måde som "subjekt" og "verbal" er grammatiske kategorier. Ligesom en sætning kan analyseres ved et grendiagram, hvor de grammatiske symboler dominerer forskellige

dele af sætningen, således kan en "indholdsstreng" også analyseres ved et grendiagram. Sætning 21 kan analyseres således som det er angivet i fig. 1.

Fig. 1.



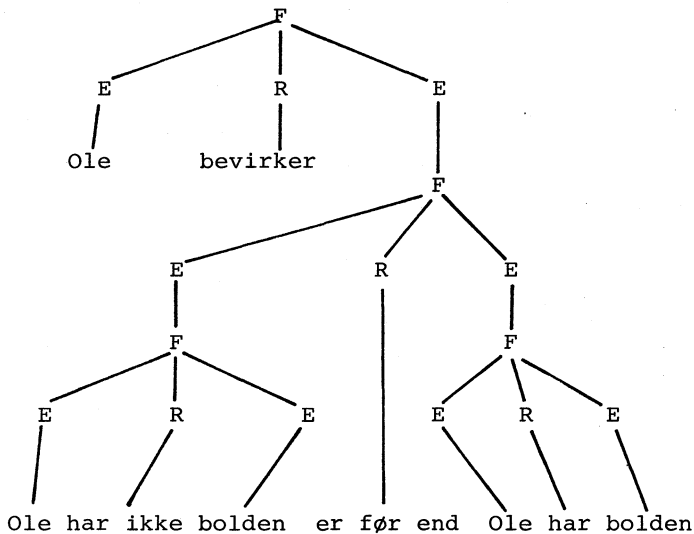
Sætning 22 og 23 er anbragt under E'erne. Det viser, at disse sætninger er de elementer, som ordnes. Under R er anbragt et element "er før end". Det betyder, at den relation, som ordner elementerne, er en tidsrelation. Man kan med lige så stor ret opfatte sætning 22 og 23 som forhold. I sætning 23 er relationen en "have-relation" mens "Ole" og "bolden" er de elementer, som ordnes af denne relation. Man ser derved, at et forhold (som sætning 23) selv kan indgå i et nyt forhold, hvor det fungerer som element. Det er nu let at notere den kausativskomponent, som fandtes i sætning 21: det var Ole - og ikke nogen anden - som bevirkede, at han kom til at have bolden. Man kan sige, at der mellem Ole og den forandring, som finder sted (nemlig at Ole først ikke har bolden og at han derefter har den) hersker en "årsagsrelation". Hele (21) kan derefter repræsenteres som angivet i fig. 2.

Diagrammet skal forklare de regelmæssigheder, man kan finde ved at foretage prøver som (16) og (18). Det vil sige, at skønt diagrammet er en abstraktion, så har det dog "empiriske" konsekvenser. Man kan kontrollere om et ords semantiske struktur er opbygget som diagrammet viser. Når diagrammet indeholder en "årsagsrelation", så skal (19) foretrækkes fremfor (20); hvis ikke skal man foretrække (20) fremfor (19). Man kan fastslå at

(25) Ole blev syg.

ikke indeholder en kausativskomponent, fordi

Fig. 2.



(26) Hvad foretog Ole sig så? Han blev syg.

føles mærkelig og afvigende, mens

(27) Hvad skete der så med Ole? Han blev syg.

er helt normal. I Bendix (66), hvorpå det foregående bygger, gives flere eksempler på sådanne tests.

Rimeligvis er "årsags-relationen" og tidsrelationen vigtige relationer; i dansk er der udviklet særlige ord til at angive, at disse relationer hersker mellem en hovedsætning og en bisætning eller mellem to hovedsætninger:

(28) Fordi Ole slog Søren, fik han bolden.

Ole slog Søren, så at han fik bolden.

(29) Efter at Ole havde slået Søren, fik han bolden.

Ole slog Søren, og så fik han bolden.

Man kan nævne, at netop tidsrelationen er det første periodeorgani-

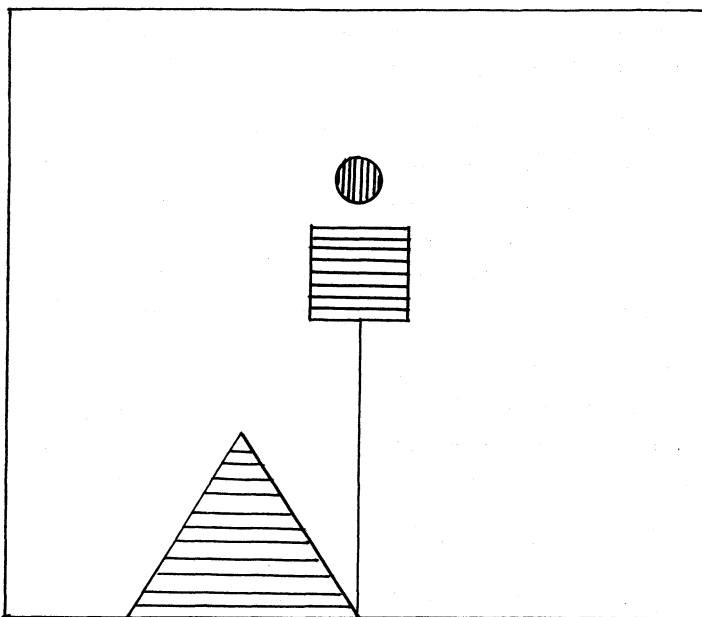


Fig. 3.

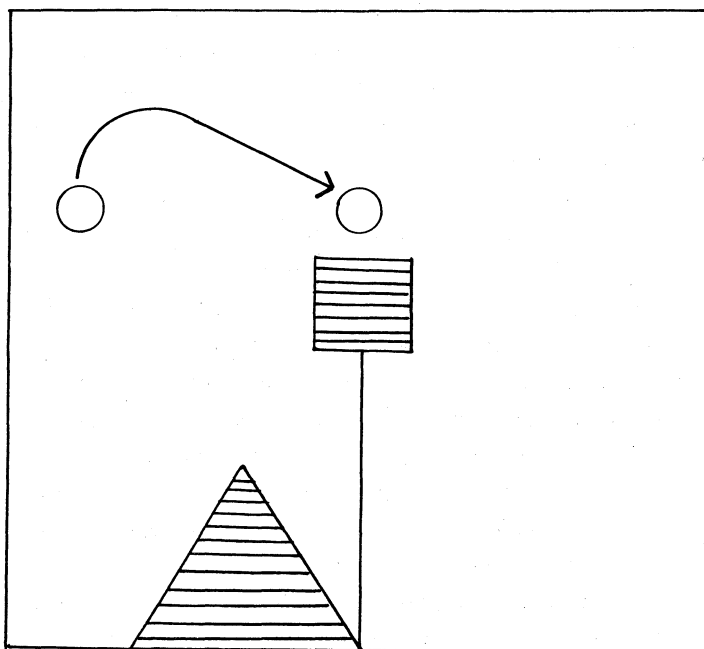


Fig. 4.

Analysen bestod nu i at arrangere disse primitive udtryk i en semantisk struktur, som den der før er omtalt. Fig. 3 kan f. eks. beskrives ved følgende primitive udtryk:

- (32) Kuglen er højere oppe end kassen.
 Kuglen er hverken til venstre eller til højre for kassen.
 Kuglen berører ikke kassen.
 Kuglen er nær ved kassen.
 Kuglen er ikke nær ved takken.
 Kuglen og kassen er højere oppe end takken.

Det viste sig, at der gjaldt visse regelmæssigheder med hensyn til på den ene side hvilke primitive relationer, der karakteriserede en tegning, og på den anden side hvilke adverbier og præpositioner, man kunne anvende. Fig. 3 kan beskrives ved en sætning som

- (33) Kuglen er oppe over kassen.

I alle de tegninger, hvor præpositionen "over" kunne anvendes, skulle følgende relationer holde mellem kuglen og kassen: kuglen skulle være højere oppe end, være hverken til venstre eller til højre for, være nær ved men ikke være i berøring med kassen. Derfor er alle disse primitive relationer komponenter af præpositionen "over". Derimod var forholdet mellem kugle/kasse og takken ikke relevant. Dette forhold betød imidlertid noget med hensyn til anvendelsen af adverbiet "oppe". "oppe" krævede, at kuglen og kassen skulle befinde sig højere oppe end et tredje objekt, takken. For at to objekter overhovedet kunne optræde på subjektets og præpositionsstyrelsens plads i de nævnte sætninger, krævedes der, at de var placeret så tæt ved hinanden, at de opfattedes som en helhed. Man kan overbevise sig om dette ved at flytte rundt med objekterne i fig. 3 og samtidig notere sig hvilke sætninger, der kan beskrive de forskellige kombinationer. Sætningen "Kuglen er oppe på kassen" kan ikke beskrive fig. 3. Det viste sig, at denne sætning kun kunne beskrive tegninger, hvor kuglen berørte kassen, mens det ikke betød noget hvilke relationer man iøvrigt opfattede i tegningen. Derfor bliver relationen "Kuglen berører kassen" komponent i præpositionen "på": et ords betydning regnes altså for at være de relationer, som nødvendigvis skal kunne opfattes i en tegning, for at en sætning, hvori ordet forekom-

mer, kan beskrive tegningen. Fig. 4 kan beskrives ved følgende sætninger:

- (34) Kuglen kom hen over kassen.
- (35) Kuglen kom ned over kassen.
- (36) Kuglen kom hen til kassen.
- (37) Kuglen kom ned til kassen.

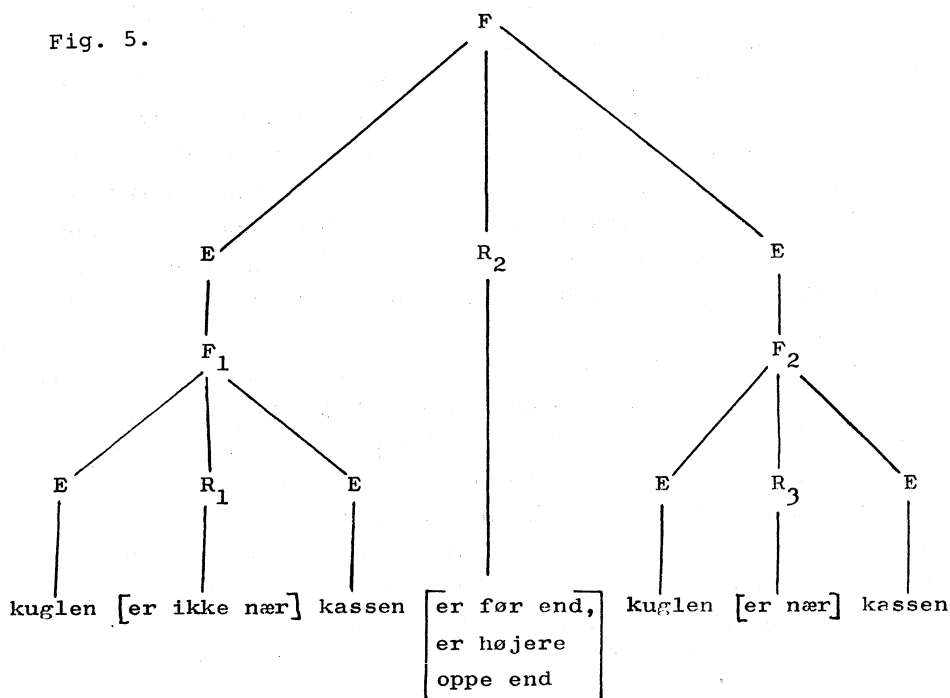
I tegningen er både kuglens slutposition og dens begyndelsesposition indtegnet. Slutpositionen er den samme som den, der er indtegnet i fig. 3. For præpositionen "til" viste det sig, at kuglens slutposition, i de tegninger, som en sætning med denne præposition kunne beskrive, skulle være nær ved kassen, men ikke nødvendigvis i berøring med kassen. Endvidere skulle der findes en position, som var før slutpositionen. I denne position skulle kuglen være fjernere fra kassen end den var i slutpositionen. Med præpositionen "over" forholder det sig anderledes; når denne præposition indgik i et retningsadverbial, skulle der i alle tegninger, som den kunne beskrive, findes en position som den, der er indtegnet i fig. 3. Men denne position skulle ikke nødvendigvis være slutpositionen, således som præpositionen "til" krævede. Man kan overbevise sig om dette ved at forlænge kuglens bane til højre: nu kan "Kuglen kom ned til kassen" ikke længere anvendes, men det kan "Kuglen kom ned over kassen" stadig. Vender man pilen om, kan ingen af sætningerne anvendes; det kan derimod sætningen "Kuglen kom op fra kassen". Præpositionen kræver noget med hensyn til kuglens begyndelsesposition, nemlig at kuglen skal være nær ved kassen samt at der skal være en position, som ligger efter begyndelsespositionen og hvor kuglen er fjernere fra kassen.

Det viste sig, at det var ganske bestemte positioner, der afgjorde, hvilke sætninger man kunne anvende. Præpositionerne "til" og "på" stillede således krav til slutpositionen, mens præpositionen "fra" stillede krav til begyndelsespositionen. Præpositioner som "over", "under" og "forbi" stillede ikke specielt krav til slut- eller begyndelsespositionerne. I det højeste var der 3 sådanne "distinktive" positioner. Adverbier som "op", "ned" og "hen" stillede krav til, hvorledes forholdet mellem de distinktive positioner var opfattet. "op" krævede således, at den første distinktive position skulle være lave-

re nede end den anden, mens "hen" krævede, at de skulle ligge til højre eller til venstre for hinanden. Fig. 4 viser, at det netop er relationerne mellem de distinktive positioner, adverbierne beskriver. Man kan beskrive tegningen ved "Kuglen kom ned til kassen" men ikke ved "Kuglen kom op til kassen". Nu findes der ganske vist to positioner, hvoraf den ene er før end og lavere nede end den anden, nemlig begyndelsespositionen og positionen i toppen af kurven. Man skulle derfor tro, at adverbiet "op" kunne anvendes her. At det ikke er tilfældet, forklares ved, at disse to positioner ikke er de distinktive for præpositionen "til". For denne præposition opfattes slutpositionen og en position, som ligger før (rimeligvis den "i toppen" af kurven), som distinktive. Det er relationen mellem disse, adverbiet beskriver.

De distinktive positioner er relationerne mellem kuglen og kassen på forskellige tidspunkter. Derfor kan betydningen af præpositioner som "til", "forbi" osv. beskrives som to eller flere forhold (F'er), som er ordnet i tid. Sætningen "Kuglen kom ned til kassen" kan repræsenteres som det er vist i fig. 5.

Fig. 5.



Diagrammet viser, at de to distinktive positioner (= de to yderste F'er) er dobbelt ordnede: den ene er før end den anden og højere oppe end den anden. Diagrammet aflæses altså således: "F₁ er før end F₂" og "F₁ er højere oppe end F₂". Man kan opfatte den kantede parentes som en konjunktion af to relationer. En anden tolkning er denne: relationerne "er før end" og "er højere oppe end" kan opfattes som karakteristika ved relationen R₂. De er "træk" eller "egenskaber", som relationen besidder. Ligegyldigt hvilken tolkning man vælger, repræsenterer diagrammet, at de elementer, som står i den kantede parentes, er ordnede med hensyn til det, der kommer før, og det, der kommer efter. Det er f.eks. forkert at bytte "er højere oppe end" om med "er ikke nær"; derimod er relationerne "er før end" og "er højere oppe end" ikke ordnede: det er ligegyldigt hvilken man skriver først.

En lignende notation anvendes i fonologien. Man kan opfatte de enkelte sproglyd som en samling af fonologiske træk. Ordet "du" kunne f.eks. repræsenteres således:

d	u
$\left[\begin{array}{l} + \text{klusil} \\ + \text{aspireret} \\ + \text{alveolær} \end{array} \right]$	$\left[\begin{array}{l} + \text{snæver} \\ + \text{bagtunge} \\ + \text{vokal} \end{array} \right]$

Også her er det ligegyldigt i hvilken rækkefølge man opskriver de fonologiske træk, som står i samme parentes. Derimod er ethvert træk i én parentes ordnet i forhold til ethvert træk i en anden parentes; det er f.eks. forkert at bytte +klusil og +snæver om, mens det er korrekt at bytte +klusil og +aspireret om. Det foregående skulle have vist at sådanne "stepmatricer" også kan anvendes i semantisk beskrivelse.

Analysen, som er blevet ridset op, er et eksempel på at sproget ligesom inddeler den oplevede virkelighed. Man oplever et objekts bevægelse som kontinuerlig, men man beskriver den ved bestemte positioner, som objektet indtager. Man inddeler bevægelsen i bestemte "faser".

Et tredje sprogligt fænomen, der med fordel kan opfattes som semantisk, er ordenes distribution. En generativ grammatik skal karakterisere en harpe således, at de og kun de strenge, som hører

til harpen, bliver genereret af systemet. Alle strenge som føles anormale skal udelukkes fra harpen. Man må således udelukke følgende frase fra den harpe, som udgøres af det danske udtrykssystem:

(38) det smilende hus

fordi udtrykket føles afvigende. Dette kan lade sig gøre ved kontekstsensitive regler. En kontekstsensitiv regel er en regel, som siger, at et ord kun kan forekomme i bestemte kontekster. Adjektivet "smilende" må således kun forekomme foran et substantiv, som betegner et menneske. I nogle tilfælde kan man rimeligvis opstille generelle regler, men i andre tilfælde modsætter sproget sig enhver form for systematisering. Ofte kan tilstandsverber ikke forbindes med adverbialer, som betegner retning. Således kan man ikke have:

(39) Han sørger op forbi skoven.

Han sover op forbi skoven.

Fra denne regel findes imidlertid utallige undtagelser. Således kan man udmærket have:

(40) Huset lå op til skoven.

Huset lå op mod skoven.

Han lå hen over bordet.

Han lå ud af vinduet.

Nogle præpositioner kan ikke forbindes med adverbierne "oppe", "nede" og "henne". Man kan ikke have:

(41) Han ligger oppe fra skoven.

Han ligger oppe til skoven.

Han løber oppe til skoven.

Men hvorfor kan man da have:

(42) Han stammer oppe fra skoven.

Han er nede fra byen.

I disse tilfælde ville der næsten kræves en kontekstsensitiv regel for enhver ordkombination; det indicerer, at den generative grammatik

har angrebet denne sag på en fundamentalt forkert måde. Ligesom ved implikationsforholdene er der tale om forhold, der er idiosynkratiske for hvert enkelt ordpar, og det taler for, at man hellere skal opfatte reglerne for ordkombinationen som semantisk betingede. Disse spørgsmål kan rimeligvis behandles mere tilfredsstillende, når man regner med et selvstændigt semantisk system "bagved" udtryks-systemet. Et eksempel vil vise hvorledes.

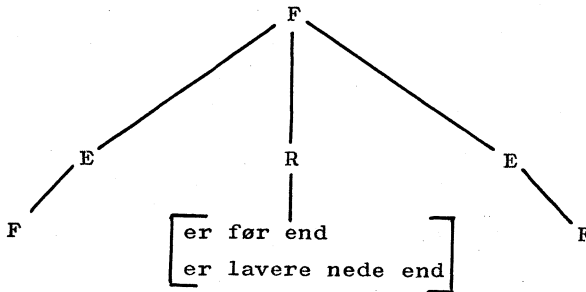
Man kan have

(43) Han faldt ned fra grenen.

men vanskeligere

(44) Han faldt op fra grenen.

Fig. 6.

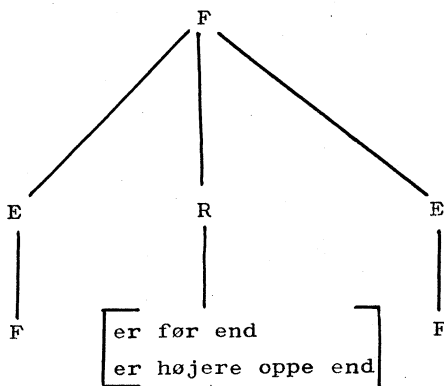


Som før nævnt krævede adverbiet "op", at forholdet mellem de distinktive positioner skulle være karakteriseret ved "er før end" og "er lavere nede end". Det vil sige, at indholdsstrukturen, som knyttes til en sætning, hvori "op" indgår, skal indeholde den konfiguration, som er vist i fig. 6. "Ordbog over det danske Sprog" definerer verbet "falde" således:

- 1) bevæge sig nedad if. tyngdeloven.
- 1.1) om det ikke understøttede legemes (hurtige) bevægelse nedad..., ofte m. særligt henblik på bevægelsernes afslutning (komme ell. kastes til jorden o.lgn.).
- 2) fra opret stilling bevæge sig nedad til liggende ell. sammenfalden stilling.

Man kan opfatte denne definition som en uformel beskrivelse af den indholdsstruktur, som udgør verbets betydning. I det system, hvori der arbejdes her, kan man oversætte det til, at der skal være en position, som er før end og som er højere oppe end en anden position ("fra..stilling..nedad til...stilling"). Verbet "falde" stiller da den betingelse til indholdsstrukturen, der knyttes til en sætning, hvori verbet forekommer, at den skal indeholde den konfiguration, der er vist i fig. 7.

Fig. 7.



Det er nu klart, hvorfor man vanskeligt kan kombinere verbet "falde" og adverbiet "op"; verbet kræver, at R skal være karakteriseret ved trækket "er højere oppe end", mens adverbiet kræver, at R skal være karakteriseret ved "er lavere nede end". Men noget kan ikke være lavere nede end noget andet og samtidig være højere oppe end det. De primitive relationer "er lavere nede end" og "er højere oppe end" kan med andre ord ikke forekomme i samme matrix (på samme måde som det fonologiske træk "ikke-aspireret" ikke kan stå i samme matrix som "aspireret"), men dette kræves netop af sætning 44. Der findes ingen indholdsstreng, som tilfredsstiller de betingelser, der stilles af ordene i sætningen, og sætningen er derfor afvigende således, som den intuitivt set skulle være. På en lignende måde vil det måske være muligt at gøre rede for eksemplerne 40-43.

Lad mig opsummere: udtryksharpen i sproget frembringes ved at udtrykssystemet genererer en base, som derefter undergår forskellige transformationer. Indholdsharpen dannes også af et generativt sy-

stem. Den grundlæggende konfiguration er forholdet F. Indholdsharpen består som udtryksharpen af strenge, der har en bestemt struktur.

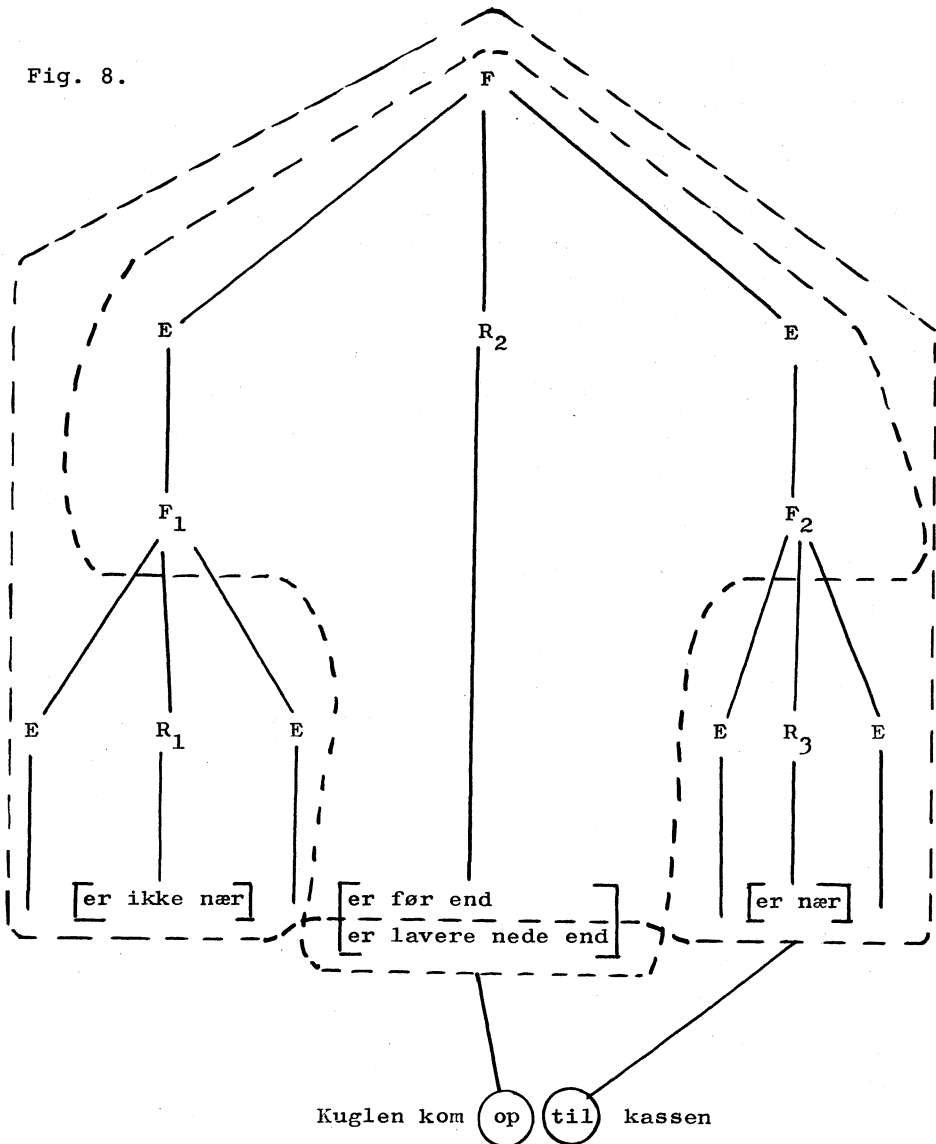
De to harper er forbundet ved en relation T, som består af en finit mængde af ordnede par. Venstreelementet i disse par er en lydform, som er betydningsbærende, et morfem eller et ord. Rime- ligvis skal oplysning om ordets syntaktiske position indgå. Højreele- mentet i relationen består af indholdsstørrelser, idet højresiden an- giver hvilke betingelser en indholdsstreng skal tilfredsstille for at kunne fungere som betydning af den sætning, hvori lydformen indgår (skellet mellem "betydning" og "referent" introduceres senere). Man kan jävnføre med begrebet "conditions for denotation" i Morris (38). Fig. 8 viser grafisk to elementer i T. Præpositionen "til" kræver, at der skal findes to forhold i indholdsstrukturen (F_1 og F_2), samt at de skal være ordnet ved en relation R_2 , som er karakteriseret ved trækket "er før end". Der kræves imidlertid intet om, at R_2 og- så skal være karakteriseret ved "er lavere nede end". Dette kræves derimod af adverbiet "op", der stiller den betingelse, at indholds- strukturen skal indeholde to F'er, samt at disse forhold skal være ordnet ved en relation, som både er karakteriseret ved "er før end" og "er lavere nede end"; som diagrammet viser, kræver "op" intet med hensyn til R_1 og R_3 . Det gør præpositionen "til" derimod.

Man ser, at ordene både kan stille betingelser til indholdsstren- gens struktur (til de semantiske kategorier "relation" (R), "element" (E) og "forhold" (F)) og til de semantiske træk (= de primitive rela- tioner). Til en sætning knyttes den mængde af indholdsstrenge, der tilfredsstiller betingelserne, som stilles af samtlige ord/morfemer i sætningen. Hvor ordene stiller betingelser, der er uforenelige føles sætningen semantisk afvigende.

Vi har beskrevet kodningen som en simpel relation T. Denne idé vil også være bærende i det følgende, men den er ikke præcis nok. En mere "sophisticated" beskrivelse er nødvendig.

Jeg har forsøgt at vise, hvorledes betydningen af verber, ad- verbier og præpositioner kan repræsenteres ved linære strukturer. Endvidere har jeg søgt at vise, hvorledes ordenes/morfemernes be- tydning ligesom "flettes ind i hinanden", hvorved kontekstuelle re- striktioner kan forklares. Senere skal jeg vise noget tilsvarende for tempus- og supinumsmorfemer.

Fig. 8.



Vil man medgive, at disse analyser er acceptable og forklarer noget om sproget, så er de anførte eksempler vidnesbyrd om, at sprogets indholdsside er linært struktureret. Dette er en ikke-triviel, empirisk konstatering. Faktisk er andre formalismer foreslået. Wallace og Atkins (60) plæderer for uordnede mængder, Lamb (64), Hockett (66) og Hays (67) foreslår, at man bruger mere generelle grafer, nemlig endelige, forbundne grafer, som beskrivelsesmodel.

Brøndals synonymiske definitioner er også uordnede mængder (jvnf. Brøndal (28) og Brøndal (40)).^{x)}

I det følgende skal jeg diskutere selve kodningsmekanismen. Fremstillingen ovenfor var noget simplificeret; til en sætning svarede et endeligt antal indholdsstreng. Nu er det imidlertid sådan, at enhver sætning kan beskrive et uendeligt antal forskellige situationer. I det typiske tilfælde er det endvidere således, at en sprogbruger, når han præsenteres for en sætning og for en vilkårlig situation, kan afgøre om sætningen kan eller ikke kan beskrive situationen (naturligvis under forudsætning af, at han har de nødvendige oplysninger om situationen). Denne evne er en del af hans sproglige viden.

I det typiske tilfælde bestemmer en sætnings indhold således en rekursiv mængde af situationer. Her må opstilles en vigtig distinktion; man må skelne mellem den fysiske situation og en repræsentation af situationen. Vi ønsker ganske vist at skrive en kodningsprocedure, men det er absurd at skrive en kodning, der koder symboler over i fysiske foreteelser. En kodning er i formel henseende en symbolmanipulation, men det er absurd at lade kodningsreglerne manipulere fysiske objekter qva symboler. Man kan derimod godt sige, at symbolske repræsentationer for en fysisk situation manipuleres. En symbolsk repræsentation af en situation kaldes en situationsrepræsentation (SR).

Man kan måske anse nervebanernes ennervering og ikke-ennervering som den fysiske manifestation af en SR. Lingvistens viden rækker dog sjældent til at beskrive, hvorledes fysiske impulser via receptorerne omformes til nerveimpulser. Han kan derimod godt spekulere over de underliggende abstrakte systemer.

Vi kan nu genformulere problemet mere præcist: et udtryk dekodet i et indhold og dette indhold karakteriserer en endelig eller uendelig mængde af SR'er. Denne mængde indeholder alle repræsentationer og kun repræsentationer af de fysiske situationer, som sætningen kan betegne. En SR forudsættes at være en streng med en struktur.

x) De her udviklede idéer udsprang ganske vist af en praktisk analyse af adverbier og præpositioner, men jeg har senere bemærket deres slægtskab med Lambs og Brøndals idéer.

Jeg skal bruge udtrykket "referent til sætningen \underline{s} " som en forkortelse for "en repræsentation af en situation, hvortil \underline{s} kan referere". I det følgende skitseres to formuleringer af dekodningsmekanismen.

1^0 . Der findes en endelig eller uendelig mængde af SR'er. En sætning \underline{s} dekodes i en indholdsstreng \underline{i} . Der defineres en operation O på enhver SR. Lad O være "transitiv" i den forstand, at hvis $O(\underline{x}) = \underline{y}$ og $O(\underline{y}) = \underline{z}$, så $O(\underline{x}) = \underline{z}$. Hvis $O(\underline{j}) = \underline{i}$ så er \underline{j} referent til \underline{i} .^{x)}

Hvis der ikke findes noget \underline{j} sådan at $O(\underline{j}) = \underline{i}$ da har \underline{s} ingen referenter og \underline{s} er semantisk anormal.

Lad \underline{i}_1 og \underline{i}_2 være indholdet af \underline{s}_1 henholdsvis \underline{s}_2 . Lad $O(\underline{i}_2) = \underline{i}_1$.

Da vil referentmængden til \underline{s}_2 indeholde referentmængden til \underline{s}_1 og \underline{s}_1 vil implicere \underline{s}_2 . For på grund af O 's transitivitet har vi nemlig at $O(\underline{j}) = \underline{i}_2$ og $O(\underline{i}_2) = \underline{i}_1$ medfører $O(\underline{j}) = \underline{i}_1$. Dvs. at \underline{j} også er referent til \underline{i}_1 . Den operation, som implicit blev brugt under eksempel gennemgangen, er: "find et deltræ af et trædiagram!" Denne operation er netop transitiv: for hvis \underline{y} er et deltræ af \underline{x} og \underline{z} er et deltræ af \underline{y} så er \underline{z} et deltræ af \underline{x} .

Jeg ved ikke om denne operation generelt er et egnet eksplikatum for "implikation" i naturligt sprog. Der synes at være tilfælde, hvor den ikke kan bruges.

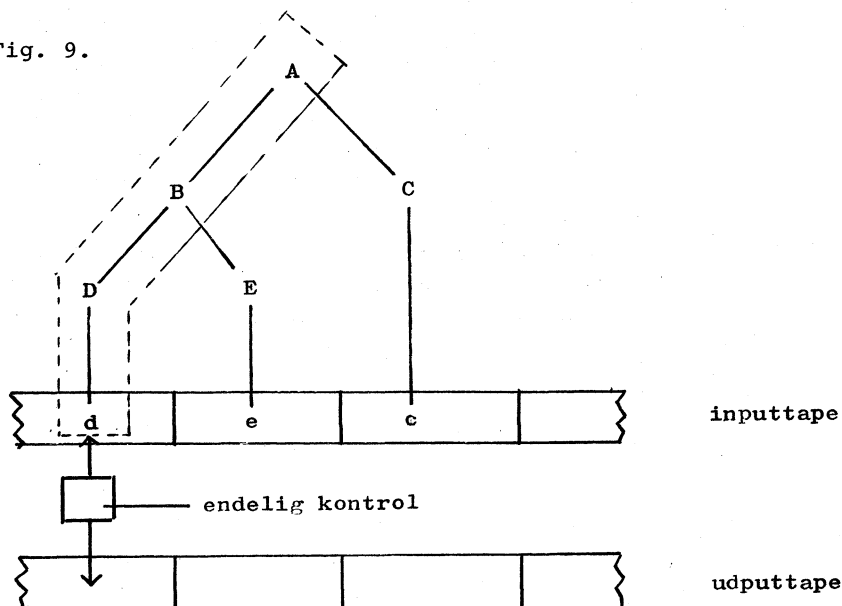
Som bekendt afhænger dekodningen af et morfem af dets syntaktiske plads. Kodningsmekanismen må da have adgang til nogle af de nonterminale symboler, der deriverer morfemet.

Hvis der findes et endeligt tal \underline{n} sådan, at antallet af nødvendige nonterminale symboler altid er mindre eller lig med \underline{n} og udtryksgrammatikken G er kontekstfri, så findes der en PDS-automat, som er ækvivalent til en automat der 1) accepterer $L(G)$ og 2) har adgang til \underline{n} symboler over ethvert terminalt symbol (jvnf. Hopcroft og Ullmann (69), theorem 12.8). Som kodningsmekanisme kan man derfor bruge en modificeret PDS-automat.

x) Beslægtede idéer findes i Bar-Hillel (64), hvor et logisk tegnsprog behandles.

Lad os tænke os kodningsmekanismen som en PDS-automat, konstrueret af G, som er i stand til at se på n symboler på sin PDS-tape og som er udstyret med et udputtape, hvorpå indholdet kan skrives. Fig. 9 søger at vise en sådan mekanisme grafisk. De stiplede linjer indeholder de symboler, som automaten kan læse.

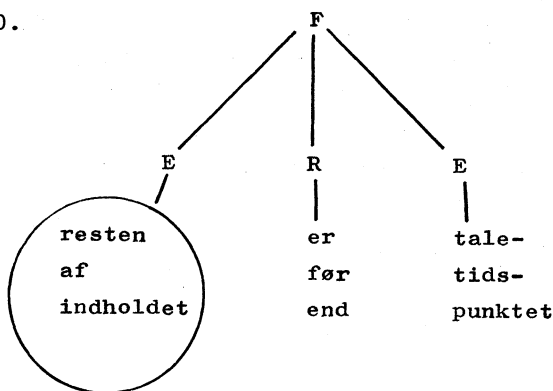
Fig. 9.



Naturligt sprog er imidlertid således indrettet, at en dekodningsmekanisme af denne art får vanskelige arbejdsforhold. Det ligger i det faktum, at udtryk og indhold ikke på nogen måde er isomorfe. Vanskeligheden opstår så snart man forsøger at skrive præcise kodningsregler. Modale og temporale udtryk modificerer ofte resten af sætningsindholdet. Hvis de sætter en relation mellem sætningsindholdet og taleren, skal denne relation stå yderst til højre eller til venstre. Lad /præteritum/ have betydningen "er før taletidspunktet". En sætning, som indeholder præteritumsmorfemet, må da have et indhold som vist i fig. 10. Lad os prøve at dekode sætningen

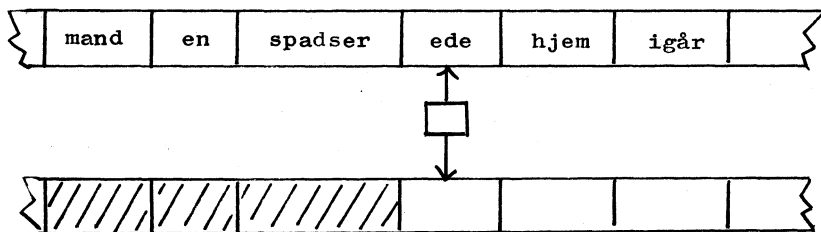
(45) /mand/en/spadser/ede/hjem/igår/

Fig. 10.



Mekanismen læser først /mand/en/spadser/ og skriver indholdet. Så skal den dekode /ede/. Situationen er vist i fig. 11. Indholdet af /ede/ skal stå yderst til højre - men mekanismen ved ikke hvor "yderst til højre" er, for den ved endnu ikke hvilke og hvormange led der kommer efter. Følgelig ved den ikke, hvormeget plads "resten af indholdet" skal optage. Det bliver da helt vilkårligt hvor indholdet af /ede/ skal skrives.

Fig. 11.



Denne vanskelighed kan afhjælpes ved at man udstyrer kodningsmekanismen med flere udputtapes og definerer en ny mekanisme, der efter at båndene er udfyldt og PDS-automaten har accepteret sætningen, løber båndene igennem og skriver det færdige indhold. Kodningsmekanismen bliver da todelt.^{x)}

Jeg skal eksemplificere brugen af disse tapes ved tempus- og supinumsmorfemerne.

Tempusmorfemerne har ikke kun temporal betydning, men jeg skal holde mig til denne her. Betydningen af supinum er strukturel.

x) Denne idé skylder jeg Brian H. Mayoh, Århus Universitets Regnecenter.

Den ligesom "ændrer" de temporale helhedsdannelser. Strukturen skal jeg angive med parenteser. Da en eksplicit kodningsanvisning er uoverskuelig skal jeg bruge "sætningsskemaprincippet". Et felt svarer til et bånd. Jeg skal kun behandle (være+ptc)-supinet.

Fig. 12 viser hvilke felter indholdet af de forskellige morfemer skal placeres i. Materialet er et udpluk fra en beskrivelse, der også omfatter tids, steds- og retningsadverbialer. Deraf de manglende numre.

Indholdsgrammatikken ser således ud:

F → E R E

F → E P

E → F

E kan desuden genskrives som repræsentationer af objekter (\approx Brøndals relatum). R kan genskrives som matricer af relationer (\approx Brøndals relator) og P kan genskrives som matricer af tilstande (\approx Brøndals descriptor).

/vågne/ betegner en overgang, dvs. to tilstande, som følger hinanden i tid. /sove/ og /vågen/ betegner tilstande.

Fig. 12.

	1	2	3	4	5	6	7	12	13	14
præsens	([+FØR]	[TT])
præteritum	([+FØR]	[TT])
være+ptc)		(()
være	()							
vågen		[VÅG]								
sove	([SØV])							
vågne	(([SØV])	[+FØR] [+RUM]	([VÅG]))			
han		[HAN]								
han		[HAN]				[HAN]				

Ifølge fig. 12 kan nedenstående sætninger dekodes til de viste indholdsstrengene:

(46) /han/præs/være/vågen/ "Han er vågen"

(46') (([HAN] [VÅG]) [+ FØR] [TT])

(47) /han/præs/vågne/ "Han vågner"

(47') ((([HAN] [SØV]) [+ FØR] [+ RUM] ([HAN] [VÅG])) [+ FØR] [TT])

(48) /han/præs/(være+ptc)/vågne/ "Han er vågnet"

(48') (([HAN] [SØV]) [+ FØR] [+ RUM] (([HAN] [VÅG]) [+ FØR] [TT]))

[TT] betyder "taletidspunktet". I (48') er nogle overflødige parenteser fjernet. Det anbefales læseren at tegne disse strenge ved hjælp af IC-diagrammer. Så er de lettere at overskue.

Man ser, at supinums betydning er denne: i den usammensatte tid (47) er det overgangen som helhed: søvn → vågen, der sættes i forhold til taletidspunktet. Supinum bevirker, at helhedsdannelsen ændres, således at "søvn" skubbes bagud og "vågen" sættes i direkte relation til taletidspunktet. Dette gælder også hvis tempus er /præt/. Man ser ligeledes, at indholdet af (46) indgår som et deltræ i indholdet af (48). Da burde sætning (48) implicere sætning (46). Det synes også at være tilfældet, da man umiddelbart vil foretrække (49a) fremfor (49b):

(49a) Han er vågnet, så han er vågen.

(49b) Han er vågnet, så han er ikke vågen.

Ifølge fig. 12 er indholdet af sætningen

(50) /han/præs/være+ptc/sove "Han er sovet"

identisk med denne indholdsstreng:

(50') ([HAN] [SØV]) ([+ FØR] [TT])

Her er igen fjernet overflødige parenteser. Man ser let, at indholdsgrammatikken ikke genererer nogen struktur, der indeholder (50')

som delstruktur. Sætningen er semantisk anormal, således som den også intuitivt er.

2^o. Den anden formulering af dekodningsmekanismen lyder således: Der findes igen en endelig eller uendelig mængde af SR'er. Indholdet af en sætning karakteriserer en endelig eller uendelig mængde af SR'er. Indholdet er således beskaffen, at man for enhver SR kan afgøre, om en vilkårlig SR tilhører eller ikke tilhører sætningens referentmængde. Indholdet kan derfor også beskrives som en grammatik der karakteriserer en endelig eller uendelig mængde af strenge.

Lad igen kodningsmekanismen være en PDS-automat af den ovenfor beskrevne type. Men på udputtapen skriver den ikke længere indholdssymboler. Den skriver symboler, der definerer en grammatik. Måske skriver den grammatiske regler.

En sprogbruger kan som før i uendelig mange tilfælde afgøre om en sætning er semantisk anormal. Det vil sige, at han kan afgøre om indholdet (som nu er en grammatik) genererer den tomme mængde af strenge eller ej. For kontekstfrie grammatikker men ikke for kontextsensitive grammatikker findes der en algoritme, hvorved dette kan afgøres (jvnf. Hopcroft og Ullmann, theorem 4.1). Derfor er det rimeligt at antage, at indholdet er kontekstfri eller simplere.

En sprogudøver kan ligeledes i uendelig mange tilfælde afgøre om en sætning implicerer en anden sætning. Han kan altså afgøre om indholdet af én sætning genererer en referentmængde, som er indeholdt i en referentmængde, der genereres af en anden sætnings indhold. Til løsning af spørgsmålet: er $L(G_1) \subseteq L(G_2)$ findes der ikke nogen generel algoritme for kontekstfrie grammatikker. Den findes for regulære grammatikker (jvnf. Hopcroft og Ullmann fig. 14.2).

Disse formelle karakteristika "beviser" ikke noget. Der siges kun, at spørgsmålene er løselige/uløselige for klassen som helhed. Spørgsmålet kan f.eks. være løseligt for nogle men ikke for andre grammatikker i klassen; alligevel siges problemet at være uløseligt for klassen som helhed.

Dog kan disse ting, sammenholdt med, at SR'er har en strukturel beskrivelse, antyde at indholdsgrammatikken befinder sig mellem de kontekstfrie eller simplere grammatikker; for kontextsensitive gram-

matikker findes der nemlig ikke nogen generel, ikke-arbitrær procedure, hvorved man kan konstruere trædiagrammer.

Den idé; at analysere og afgrænse klasser af fysiske objekter ved hjælp af tilsvarende grammatikker er i de senere år vundet frem indenfor "computer science". Shaw (69) opstiller en grammatik, hvorved man kan analysere og afgrænse klasser af billeder. Denne grammatik er kontekstfri.

Vi må forestille os, at indholdssystemet er vehikel for menneskets tankevirksomhed. Ligesom et programmeringssprog indeholder data og instruktioner for computeren, således repræsenterer indholdssystemet data og instruktioner for mennesket. De idag brugte programmeringssprog (ALGOL-like languages) er kontekstfrie.

Måske bruger mennesket flere 'programmeringssprog'. Måske oversættes indholdssystemet til et 'nervesprog', ligesom maskinen har compilere, som oversætter programmeringssproget til maskinsproget.

Læseren bedes tage disse sidste gisninger, ikke med et gran men med en søkfuld salt.

Post Scriptum.

Idéen i artiklen var at indholdssystemet lod sig opfatte som en harpe, der kunne karakteriseres af en grammatik. Efter at have afsluttet artikeln er jeg blevet opmærksom på at andre forfattere har udtrykt lignende anskuelser. Idéerne i H.E. Brekles artikel 'Generative semantics vs deep structure' (Studies in syntax and semantics, 1969) er omtrent identiske med mine. Mekanismer, der oversætter fra syntaktisk til logisk repræsentation, bruges i mange 'question-answering'-programmer; en artikel af R.F. Simmons 'Natural language question-answering systems: 1969' (Communications of the ACM, 1970, pp. 15-30) giver en fyldig oversigt over litteraturen. R.W. Langacker lader i 'Mirror image rules II' (Language, 1969, pp. 844-863) overgangen fra semantisk til syntaktisk repræsentation ske ved transformationsregler. R. Quillian og B. Raphael antyder i 'Semantic information processing' (1968) at de semantiske repræsentationer er struktureret som grafer og ikke som strenge (jvnf. også Hockett (66)). I så fald skal indholdsgrammatikken ikke være en 'strenggrammatik' men en 'grafigrammatik' af den type som beskrives af U.G. Montena-

ri i 'Separable graphs, planar graphs and web grammars' (Information and control, 1970, pp. 243-267). Sluttelig kan jeg henvise til artikelsamlingen 'Universals in linguistiv theory' (1968) hvor flere af forfatterne diskuterer en generativ semantik. Idéerne her er dog på væsentlige punkter forskellige fra de ovenfor anførte; Den væsentligste forskel er, at disse forfattere opgiver basebegrebet, mens basen er bibeholdt i min model.

(PBA, 28-12-70)

Referencer:

- Bar-Hillel, Y.: Language and Information (Mass. 1964).
- Bendix, E.H.: Componential Analysis of general Vocabulary (International Journal of American Linguistics, vol. 32, no. 2, 1966).
- Brøndal, V.: Ordklasserne (Kbh., 1928).
Præpositionernes Theori (Kbh., 1940).
- Chomsky, N.: Formal Properties of Grammars (Handbook of Mathematical Psychology, vol. 2, New York, 1963).
- Hays, D.G.: Introduction to Computational Linguistics (New York, 1967).
- Hockett, C.F.: Language, Mathematics and Linguistics (Current Trends in Linguistics, vol. 3, The Hague, 1966).
- Hopcroft, J.E. og Ullman, J.D.: Formal Languages and their Relation to Automata (Mass., 1969).
- Koffka, K.: Principles of Gestalt Psychology (New York, 1935).
- Lamb, S.M.: The Semic Approach to Structural Semantics (Trans-cultural Studies, 1964).
- Morris, C.W.: Foundations of the Theory of Signs (Chicago, 1938, genoptrykt 1947).
- Shaw, A.C.: A Formal Picture Description Scheme as a Basis for Picture Processing Systems (Information and Control, vol. 14, no. 1, 1969).
- Wallace, A.F.C. og Atkins, J.: The meaning of kinship terms (American Anthropologist, vol. 62, 1960).